

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-306407

(43)Date of publication of application : 02.11.2001

(51)Int.Cl.

G06F 12/16

G06F 3/06

G06F 13/00

G11B 20/10

(21)Application number : 2000-118012

(22)Date of filing : 19.04.2000

(71)Applicant : HITACHI LTD

(72)Inventor : AJIMATSU YASUYUKI

MATSUNAMI NAOTO

MURAOKA KENJI

OEDA TAKASHI

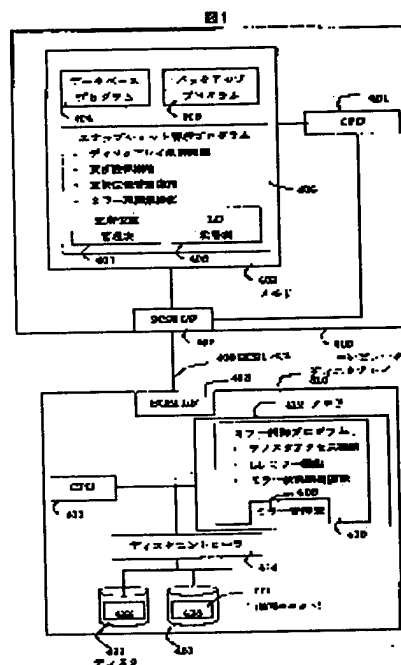
YAGISAWA IKUYA

(54) METHOD FOR MANAGING SNAPSHOT AND COMPUTER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To manage a snapshot without increasing the CPU load of a computer and the communication quantity of data due to the duplex processing of data writing by evading the concentration of load onto a specific medium in backup operation.

SOLUTION: When a snapshot is not taken, data writing from the computer to an external storage device 410 is duplicated and stored in respective areas of plural different storage media 420, 421. The duplication of data writing is performed by the external storage device 410. In the ease of taking a snapshot of external storage, one area out of the duplicated areas is provided as a storage area for a normal access and the other area is provided as a snapshot area. When the normal access storage area is updated during the storage of the snapshot, the computer allows the contents of the two areas to coincide with each other at the time of deleting the snapshot.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

W1108

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-306407

(P2001-306407A)

(43) 公開日 平成13年11月2日 (2001.11.2)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 0 6 F 12/16	3 1 0	G 0 6 F 12/16	3 1 0 M 5 B 0 1 8
			3 1 0 J 5 B 0 6 5
3/06	3 0 4	3/06	3 0 4 F 5 B 0 8 3
	5 4 0		5 4 0 5 D 0 4 4
13/00	3 0 1	13/00	3 0 1 P

審査請求 未請求 請求項の数 7 O L (全 17 頁) 最終頁に続く

(21) 出願番号 特願2000-118012 (P2000-118012)

(22) 出願日 平成12年4月19日 (2000. 4. 19)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 味松 康行

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72) 発明者 松並 直人

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74) 代理人 100078134

弁理士 武 顕次郎

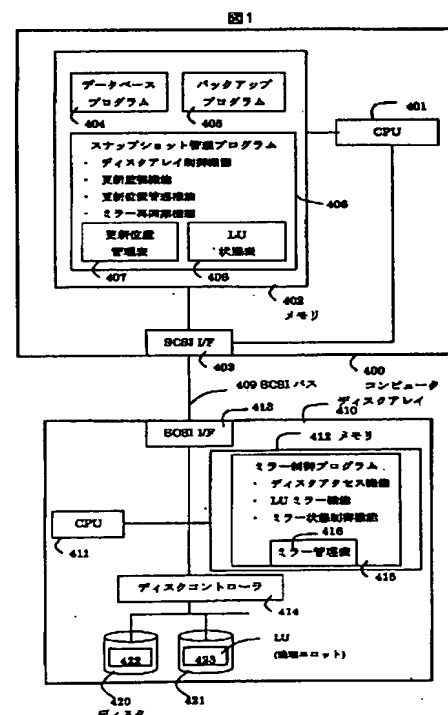
最終頁に続く

(54) 【発明の名称】 スナップショット管理方法及び計算機システム

(57) 【要約】

【課題】 バックアップ中の特定媒体への負荷の集中を避け、データ書き込みの二重化処理に伴うコンピュータのCPU負荷やデータ通信量の増加を生じさせることなくスナップショットの管理を行う。

【解決手段】 スナップショットを取っていないときには、コンピュータから外部記憶装置へのデータ書き込みを、異なる複数の記憶媒体420、421上の領域に二重化して記憶する。データ書き込みの二重化は、外部記憶装置410で行う。外部記憶のスナップショットを取る際、二重化された領域のうち1つの領域を通常のアクセス用の記憶領域として提供し、もう1つの領域をスナップショットとして提供する。スナップショット保存中に通常のアクセス用の記憶領域が更新された場合、スナップショット削除時にコンピュータが、2つの領域の内容を一致させる。



【特許請求の範囲】

【請求項 1】 複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおけるスナップショット管理方法において、前記外部記憶装置は、内部の異なる 2 つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記 2 つの記憶領域をグループとして定義し、グループ内の 2 つの記憶領域をアクセス可能な 1 つの仮想的な記憶領域としてコンピュータに提供し、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の 2 つの記憶領域に書き込んで二重化し、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記 2 つの記憶領域を独立した記憶領域としてアクセス可能とし、前記コンピュータ内で動作するプログラムは、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出し、データの書き込み先の位置を記録することを特徴とするスナップショット管理方法。

【請求項 2】 前記外部記憶装置は、各記憶領域についてその状態を記憶し、コンピュータからの指示により指定された記憶領域の状態を変更し、記憶領域の状態に応じて、記憶領域へのアクセスの制限、グループ内の 2 つの記憶領域へのデータ書き込みの二重化の制御を行い、コンピュータ内で動作するプログラムは、記憶領域の状態を記憶し、外部記憶装置に記憶領域の状態を変更する指示を送信し、前記外部記憶装置は、スナップショット削除時に、記録された更新位置のデータを記憶領域間でコピーすることによりグループ内の 2 つの記憶領域の内容を一致させることを特徴とする請求項 1 記載のスナップショット管理方法。

【請求項 3】 スナップショット削除時、コンピュータ内で動作するプログラムは、コピーする必要があるデータの位置を外部記憶装置に送信し、外部記憶装置は、その位置情報に基づいて領域間でデータコピーを行うことを特徴とする請求項 2 記載のスナップショット管理方法。

【請求項 4】 コンピュータ内で動作するプログラムは、記憶領域の状態を一括管理し、スナップショット削除時に、記録された更新位置のデータを記憶領域間でコピーすることによりグループ内の 2 つの記憶領域の内容を一致させることを特徴とする請求項 1 記載のスナップショット管理方法。

【請求項 5】 ネットワークで接続された別のコンピュータ上のプログラムが、スナップショットの取得及び削除を指示し、外部記憶装置が、オリジナルのデータにアクセスするコンピュータとは異なるコンピュータにスナップショットを提供することを特徴とする請求項 4 記載のスナップショット管理方法。

【請求項 6】 コンピュータ内で動作するプログラムは、スナップショット取得後に、スナップショットを取

得された記憶領域の内容を更新する際、更新前のデータを別の記憶領域に保存すると共に、保存先の位置を記録し、スナップショットを、コンピュータ上で動作する他のプログラムがアクセスできる仮想的な記憶領域として提供し、他のプログラムがスナップショットの読み出しを要求したとき、アクセス先がスナップショット取得後に更新されていれば、別の記憶領域に保存された更新前のデータを読み出し、アクセス先が更新されていなければ、外部記憶装置が提供する仮想的な記憶領域を構成する 2 つの記憶領域のうち、特定の 1 つからデータを読み出すことを特徴とする請求項 1 記載のスナップショット管理方法。

【請求項 7】 複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおいて、前記外部記憶装置は、内部の異なる 2 つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記 2 つの記憶領域をグループとして定義し、グループ内の 2 つの記憶領域をアクセス可能な 1 つの仮想的な記憶領域としてコンピュータに提供する手段と、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の 2 つの記憶領域に書き込んで二重化する手段と、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記 2 つの記憶領域を独立した記憶領域としてアクセス可能とする手段とを備え、前記コンピュータは、その内部で動作するプログラムが、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出する手段と、データの書き込み先の位置を記録する手段とを備えることを特徴とする計算機システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、スナップショット管理方法及び計算機システムに係り、特に、高可用性を求められる計算機システムにおける外部記憶をバックアップする際に必要な外部記憶のスナップショット管理方法、及び、この管理方法により管理される記憶装置を有する計算機システムに関する。

【0002】

【従来の技術】一般に、ハードディスク等のコンピュータの外部記憶装置に記録されたデータは、装置の障害、ソフトウェアの欠陥、誤操作などによりデータを喪失した場合に、喪失したデータを回復できるように定期的に磁気テープ等にコピーして保存しておくこと（バックアップ）が必要である。その際、コピー作業中にデータが更新されるとコピーしたデータに不整合が生じるため、コピー作業中にデータが更新されないことを保証する必要がある。

【0003】バックアップされるデータの更新を避けるためには、データにアクセスするバックアッププログラ

ム以外のプログラムを停止させればよいが、高可用性が要求されるシステムは、プログラムを長時間停止させることができない。このため、バックアップ中にプログラムがデータを更新することを妨げず、なおかつバックアップ開始時点でのデータの記憶イメージを保存する仕組みが必要である。

【0004】通常、ある時点でのデータの記憶イメージをスナップショットと呼び、指定された時点のスナップショットを保存しつつデータの更新が可能な状態を提供する仕組みをスナップショット管理方法と呼ぶ。また、スナップショット管理方法によりスナップショットを保存することをスナップショットの取得と呼び、スナップショット取得の対象となったデータをオリジナルデータと呼ぶ。

【0005】従来のスナップショット管理方法は、更新前データの保存、コンピュータによるデータの二重化、あるいは、外部記憶装置によるデータ二重化によって実現されている。以下、これらの方法による従来のスナップショット管理方法について説明する。

【0006】(1) 更新前データの保存による方法
この方法に関する従来技術として、例えば、米国特許第5649152号明細書に記載された技術が知られている。この従来技術は、あるデータのスナップショットを取得するとき、スナップショット取得時点以後にオリジナルデータを更新する場合、更新前の記憶内容を別の記憶領域に保存するというものである。スナップショットは、論理的にはオリジナルデータとは独立な別データとしてアクセスされるが、スナップショット取得時点以後にオリジナルデータが更新されていない部分について、スナップショットはオリジナルデータと記憶領域を共有する。オリジナルデータが更新された部分は、別領域に保存された更新前の記憶内容が参照される。また、この場合、スナップショットは読み出し専用である。

【0007】(2) コンピュータによるデータの二重化による方法

コンピュータによるデータの二重化による方法は、スナップショットを取得していない通常の状態において、コンピュータ上のプログラムが全てのデータを2つの記憶領域に二重化（ミラー）するというものである。スナップショットを取得するとき、コンピュータ上のプログラムは、二重化の処理を停止して2つの記憶領域を独立な領域に分離し、1つの領域をオリジナルデータ、もう1つの領域をスナップショットとして提供する。コンピュータ上のプログラムがデータを二重化する方法は、例えば、米国特許5051887号明細書等に記載されて知られている。

【0008】(3) 外部記憶装置によるデータの二重化による方法

外部記憶装置によるデータの二重化による方法は、例えば、米国特許5845295号明細書に示されているよ

うに、前述した(2)の方法において、コンピュータ内のプログラムが行ったスナップショット管理を外部記憶装置内で行うというものである。スナップショットを管理する機能は、全て外部記憶装置内のスナップショット管理プログラムにある。

【0009】

【発明が解決しようとする課題】 前述した更新前データの保存による従来技術の方法は、スナップショット取得後、オリジナルデータが更新されていない部分について、スナップショットに対するアクセスも、オリジナルデータに対するアクセスも同一の記憶領域を参照する。このため、この従来技術は、バックアップ中、特定の記憶媒体にアクセスが集中し、ディスク入出力の性能が低下するという問題点を有している。この結果、例えば、計算機システムの特定のディスク装置にアクセスが集中し、データベースプログラムやバックアッププログラムがそのディスク装置のLU内のデータを読み出す速度が低下してしまうことになる。

【0010】また、この従来技術は、オリジナルデータへのアクセスとスナップショットへのアクセスとのどちらも必ずスナップショット管理プログラムを経由しなければならないため、バックアップ中、スナップショット管理プログラムを実行するコンピュータの負荷が増大し、同一コンピュータ内で動作するデータベースプログラム等の実行速度が低下するという問題点を有している。この結果、特に、大量のデータをバックアップする際等に、バックアップ終了までの長時間にわたり性能を低下させてしまうことになる。

【0011】また、前述したコンピュータによるデータの二重化による従来技術の方法は、スナップショットを取得していない状態でのデータ書き込みにおいて、1回のデータ書き込みに対してスナップショット管理プログラムが2つのLUにデータを書き込む必要があるため、スナップショット管理をしないシステムと比較して、2倍の書き込み処理を実行する必要がある。従って、この従来技術は、データの書き込みを行うコンピュータのCPU負荷や、コンピュータと外部記憶装置とを接続する通信路のデータ通信量、外部記憶装置のディスクコントローラ負荷が増大し、スナップショット管理をしないシステムと比較してアプリケーションプログラムの実行速度を低下させてしまうという問題点を生じる。特に、データベースのレプリケーション等、大量のデータ更新が伴う処理において性能の低下が顕著となる。

【0012】さらに、外部記憶装置によるデータの二重化による従来技術の方法は、スナップショット管理に必要な全ての処理を外部記憶装置内に実装するため、外部記憶装置の制御プログラムが複雑になり、複雑化に対応して開発期間が長くなり、価格も上昇するため、このような外部記憶装置を用いる構成のシステムは高価になってしまうという問題点を有している。

【0013】前述したように、従来技術は、バックアップ時の特定記憶媒体へのアクセス集中及びCPU負荷の増大、通常運用時のCPU、通信路及び外部記憶装置のコントローラの負荷の増大、あるいは、外部記憶装置の複雑化と高価格化という問題点を有している。

【0014】本発明の目的は、前述した従来技術の問題点を解決し、外部記憶のスナップショットを取得した状態において、外部記憶装置の特定の記憶媒体への負荷集中やコンピュータのCPU負荷の増大を防ぐことができ、外部記憶のスナップショットを取得していない状態において、コンピュータのCPU負荷やデータ通信路の通信量及び外部記憶装置の負荷の低いスナップショット管理方法を提供することにある。

【0015】また、本発明の目的は、比較的単純な機能だけを持つ安価な外部記憶装置を用いて、外部記憶のスナップショット管理を行う計算機システムを提供することにある。

【0016】

【課題を解決するための手段】本発明によれば、複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおけるスナップショット管理方法において、前記外部記憶装置が、内部の異なる2つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記2つの記憶領域をグループとして定義し、グループ内の2つの記憶領域をアクセス可能な1つの仮想的な記憶領域としてコンピュータに提供し、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の2つの記憶領域に書き込んで二重化し、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記2つの記憶領域を独立した記憶領域としてアクセス可能とし、前記コンピュータ内で動作するプログラムが、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出し、データの書き込み先の位置を記録することにより達成される。

【0017】

【発明の実施の形態】以下、本発明によるスナップショット管理方法及び計算機システムの実施形態を図面により詳細に説明する。

【0018】図1は本発明の第1の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図2はミラー管理表の構成を示す図、図3は更新位置管理表の構成を示す図、図4はMode Select及びMode Senseで使用するパラメータページの内容を説明する図、図5はスナップショット管理プログラムのデータ書き込み時の動作を説明するフローチャート、図6はコンピュータからWRITEコマンドを受信したミラー制御プログラムの動作を説明するフローチャート、図7はスナップショット管理プログラムのミラー再同期機能

の動作を説明するフローチャートである。図1～図3において、400はコンピュータ、401、411はCPU、401、412はメモリ、403、413はSCSIインタフェース、404はデータベースプログラム、405はバックアッププログラム、406はスナップショットプログラム、407は更新位置管理表、408はLU状態表、409はSCSIバス、410はディスクアレイ、414はディスクコントローラ、415はミラー制御プログラム、416はミラー管理表、420、421はディスク装置、422、423はLUである。本発明の第1の実施形態は、データベースプログラムがアクセスしている記憶領域を、データベース処理と並行してオンラインバックアップする場合に、本発明によりスナップショットを管理する例である。

【0019】図1に示す計算機システムは、コンピュータ400とディスクアレイ410とがSCSIインタフェース403、413を介してSCSIバス409により接続されて構成されている。コンピュータ400内のメモリ402には、データベースプログラム404、バックアッププログラム405、スナップショット管理プログラム406が格納され、コンピュータを制御するCPU401によって実行される。ディスクアレイ410には、ディスクコントローラ414によって制御されるディスク装置420、421が設けられ、また、メモリ412内には、ミラー制御プログラム412が格納され、CPU411によって実行される。各ディスク内の記憶領域は、SCSIの論理ユニット(LU)422、423としてコンピュータ400からアクセスされる。

【0020】データベースプログラム404は、実行中にLU422にアクセスし、また、ユーザからの指示によりディスク上のデータの更新を停止、再開する機能を持つ。バックアッププログラム405は、ユーザからの指示により、LU423からバックアップのためのデータ読み出しを行う。スナップショット管理プログラム406は、デバイスドライバやボリュームマネージャのようなミドルウェアであり、データベースプログラム404がディスクアレイに出すディスクアクセス要求の全てを自スナップショット管理プログラム406を経由してディスクアレイ410に発行する。そして、スナップショット管理プログラム406は、ディスクアレイ制御機能、更新監視機能、更新位置管理機能、ミラー再同期機能を有すると共に、更新位置管理表407、LU状態表408を有する。

【0021】ディスクアレイ410のミラー制御プログラム415は、コンピュータ400からの要求に応じてディスクコントローラ414にディスクアクセスを指示するディスクアクセス機能の外に、1つのLUに対する更新を二重化して予め指定された別のLUにも適用し、2つのLUに同一の内容を書き込むLUミラー機能を持つ。図示例の本発明の本実施形態は、LU422をLU

423に二重化する機能を有する。これらは従来のディスクアレイが持つ機能である。また、ミラー制御プログラム415は、LUミラー機能を無効化することにより、二重化された2つのLUをそれぞれアクセス可能な独立のLUとするミラー状態制御機能を有する。

【0022】二重化される各LUのミラー先の定義と現在の状態とは、ミラー管理表416に記録される。ミラー管理表416は、図2にその構成を示すように、二重化される各LUについて、そのミラー先LUと状態とが記録される。状態は、2つのLUの内容が一致している「同期」、それぞれが独立なLUとしてアクセスされる「スナップショット中」、「スナップショット中」から「同期」に移るためにLU間で一致しないデータをコピーしている「再同期中」の3種類がある。LUミラー機能は状態が「同期」、「再同期中」のLUに対して適用される。ミラー管理表416において、ミラー先LUを複数にすることにより、3重以上の多重化構成とすることもできる。

【0023】スナップショット管理プログラム406は、後述するように、SCSIのModeSense、Mode Selectコマンドを発行するディスクアレイ制御機能、スナップショット取得後にデータベースプログラムが要求するオリジナルデータの更新を検出する更新監視機能、その位置を記録する更新位置管理機能、及び、スナップショット削除時にオリジナルデータの更新部分をミラー先LUにコピーするミラー再同期機能を有している。更新監視機能及び更新位置管理機能は、状態が「スナップショット中」または「再同期中」のLUに対して、また、ミラー再同期機能は、状態が「再同期中」のLUに対してのみ適用される。

【0024】各LUの現在の状態を記録するLU状態表408は、図2に示すディスクアレイのミラー管理表416からミラー先LUの定義を省略した構成を有する。LU状態表408は、スナップショット管理プログラム406の起動時に、アクセス可能な全てのLUについてMode Senseコマンドを発行し、ディスクアレイ410から各LUの状態を得ることにより作成される。また、更新位置管理表407には、スナップショット取得後のオリジナルデータの更新位置が記録される。更新位置管理表には、図3に示すように、更新されたオリジナルデータを記憶するLU、更新された位置の最初のLBA、最後のLBAが記録される。図3に示す例の場合、LU2のLBA0~15及びLBA256~511、LU3のLBA8~31の部分がスナップショット取得後に更新されていることが示されている。更新位置管理表407の記録は、二重化のペアとして定義された2つのLU間で内容が一致しない位置を表す。

【0025】次に、前述したように構成される本発明第1の実施形態による計算機システムにおけるコンピュータからのディスクアレイの制御について説明する。

【0026】LUミラー機能の有効/無効化やミラー先LUに対するアクセスの可否等のディスクアレイのミラー制御プログラムの各LUに対する動作は、全てミラー管理表に記録されているLUの状態によって決定される。ミラー制御プログラム415は、ミラー管理表416の状態記録をスナップショット管理プログラム406が発行するSCSI Mode Selectコマンドで指定された状態に変更する。また、ミラー制御プログラム415は、ミラー管理表416の内容を要求するSCSI Mode Sense コマンドを受信すると、指定されたLUの現在の状態記録を返送する。Mode Select及びModeSenseで使用するパラメータページの内容を図4に示している。パラメータページは、2バイトからなり、先頭にはページコード(00h)が指定されており、2バイト目は指定されたLUの状態を表している。Mode Select は、「同期」、「スナップショット中」、「再同期中」に応じて、それぞれ、0、1、2を指定する。また、Mode Senseは、現在の状態を返送するが、ModeSenseコマンドで指定したLUが二重化されていなかったり、ミラー先LUとして定義されている場合、「未定義」として4が指定される。

【0027】次に、スナップショットの取得について説明する。LU422のスナップショットを取得するには、以下に説明するような処理を行う。

【0028】まず、ユーザが一時的にデータベースの更新を停止し、バックアップするデータの整合性を保証する。次に、スナップショット管理プログラムにLU422のスナップショットの取得を指示する。スナップショット管理プログラム406は、LU状態表408を参照し、LU422の状態が「同期」であることを確認するが、LU状態表にLU422の記録がない場合や、状態が「同期」でない場合、処理を中止する。

【0029】LU422の状態が「同期」である場合、スナップショット管理プログラム406はLU状態表408に記録された状態を「同期」から「スナップショット中」に変更して更新監視機能及び更新位置管理機能を有効化すると共に、ディスクアレイ410内のミラー管理表416に記録されたLU422の状態を「スナップショット中」に変更するためにMode Selectコマンドを発行する。ディスクアレイ410のミラー制御プログラム415は、Mode Selectコマンドを受信すると、ミラー管理表416に記録されたLU422の状態を「スナップショット中」に変更する。その後、ユーザは、データベースの更新を再開する。

【0030】データの読み出し、書き込みの制御として後述するように、ミラー制御プログラム415の各LUに対する動作はLUの状態により決定される。従って、この状態変更により、ミラー先LUであるLU423が独立したLUとしてコンピュータからアクセス可能となり、また、LU422に対する更新がLU423に反映

されなくなるため、バックアッププログラム 405 は、LU423 を LU422 のスナップショットとして利用することが可能となる。

【0031】次に、スナップショットの削除について説明する。LU422 のスナップショットを削除するには、以下に説明するような処理を行う。

【0032】まず、ユーザがスナップショット管理プログラム 406 に LU422 のスナップショットの削除を指示する。スナップショット管理プログラム 406 は、LU 状態表 408 を参照し、LU422 の状態が「スナップショット中」であることを確認するが、LU 状態表 408 に LU422 の記録がない場合や、状態が「スナップショット中」でない場合処理を中止する。

【0033】LU422 の状態が「スナップショット中」である場合、スナップショット管理プログラム 406 は、ディスクアレイ 410 内のミラー管理表 416 に記録された LU422 の状態を「再同期中」に変更するために Mode Select コマンドを発行する。ディスクアレイのミラー制御プログラム 415 は、Mode Select コマンドを受信すると、LU422 の状態を「再同期中」に変更する。この状態変更により、LU423 をアクセス先とするコンピュータからのアクセスが不可能となり、また、LU422 の更新を LU423 に反映する LU ミラー機能が有効化される。

【0034】次に、スナップショット管理プログラム 406 は、LU 状態表 408 に記録された LU422 の状態を「再同期中」に変更し、ミラー再同期機能を有効化する。ミラー再同期機能によるデータコピー処理の詳細については後述する。スナップショット管理プログラム 406 は、データコピー処理が完了すると、再び Mode Select コマンドを発行してディスクアレイ 410 内のミラー管理表 416 に記録された LU422 の状態を「同期」に変更し、LU 状態表 408 に記録された状態を「同期」に変更する。以上により、スナップショット管理プログラムの LU422 に対する更新監視機能、更新位置管理機能、ミラー再同期機能が無効化され、スナップショットの削除が完了する。

【0035】次に、データの読み出しの処理、すなわち、コンピュータ上のデータベースプログラム 404、バックアッププログラム 405 が、それぞれディスクアレイの LU422、423 内のデータを読み出すときの処理手順について説明する。

【0036】コンピュータ 400 内のプログラムからの入出力要求は、スナップショット管理プログラム 406 を経由するが、読み出しの場合、スナップショット管理プログラム 406 は何もせずに SCSI の READ コマンドをディスクアレイに転送する。ディスクアレイのミラー制御プログラム 415 は、READ コマンドを受信すると、指定された LU からデータを読み出し、データとステータスとをコンピュータ 400 に返送する。但

し、ミラー制御プログラム 415 は、指定された LU の状態が「同期」であれば、指定された LU とそのミラー先 LU の内容とが一致しているの、両者のいずれかから読み出すことにより負荷を分散させる。また、ミラー制御プログラム 415 は、指定された LU がミラー先 LU で、かつ、状態が「スナップショット中」でない場合、アクセスを拒否し、エラーステータスを返送する。なお、スナップショット中でないミラー先 LU へのアクセスは、書き込みの場合も含めてスナップショット管理プログラム 406 内で拒否するようにしてもよい。

【0037】次に、データの書き込みの処理、すなわち、コンピュータ上のデータベースプログラム 404、バックアッププログラム 405 が、それぞれディスクアレイの LU422、423 内の記憶内容を更新するときの処理動作を図 5 に示すフローを参照して説明する。

【0038】(1) スナップショット管理プログラム 406 は、プログラム 404、405 からの更新要求を受け取ると、LU 状態表 408 を参照し、アクセス先 LU の状態が「スナップショット中」であるか否かを確認する (ステップ 800、801)。

【0039】(2) ステップ 801 での判定が「スナップショット中」であった場合、アクセス先の位置がすでに更新位置管理表 407 に記録されているか否かを調べ、記録されていないければ、更新位置管理表 407 に位置を記録して、SCSI の WRITE コマンドを発行する。また、記録されていれば、そのまま SCSI の WRITE コマンドを発行する (ステップ 802、803、807)。

【0040】(3) ステップ 801 での判定で、LU の状態が「スナップショット中」でなかった場合、さらに「再同期中」か否かを確認し、もし「再同期中」でなければアクセス先 LU の状態は「同期」であるから、そのまま WRITE コマンドを発行する (ステップ 804、807)。

【0041】(4) ステップ 804 での判定で、「再同期中」であれば、さらにアクセス位置が更新位置管理表 407 に記録されているか否かを調べ、記録がなければ、そのまま WRITE コマンドを発行し、更新位置の記録があれば、表から該当位置の記録を削除してから WRITE コマンドを発行する (ステップ 805～807)。

【0042】前述において、更新位置管理表 407 の記録の追加・削除は、単純なエントリの追加・削除ではなく、追加の場合、同一位置を含む複数の記録が存在しないように、削除の場合、削除する位置以外の記録を削除しないように更新位置管理表 407 の修正を行う。例えば、更新位置管理表 407 が図 3 により説明したような内容を持ち、LU3 の LBA=10～11 を削除する場合、表に記録された LU3 の更新位置である LBA=8～31 を削除し、新たに LBA=8～9 及び LBA=1

2～31のエントリを追加する。

【0043】次に、コンピュータからWRITEコマンドを受信したディスクアレイのミラー制御プログラム415の動作を、図6に示すフローを参照して説明する。

【0044】(1) WRITEコマンドを受信したミラー制御プログラム415は、まず、ミラー管理表を参照してアクセス先LUIがミラー先LUIとして記録されているか否かを調べる(ステップ900、901)。

【0045】(2) ステップ901の判定で、アクセス先がミラー先LUIであれば、状態が「スナップショット中」であるか否かを調べ、スナップショット中であれば指定されたLUIにデータを書き込み、そうでなければ独立したLUIとしてアクセスできないので、書き込み失敗のステータスを返送する(ステップ905、904、906)。

【0046】(3) ステップ901の判定で、アクセス先LUIがミラー先LUIでない場合、状態が「スナップショット中」か否かを確認し、スナップショット中でなければミラー先LUIと指定されたLUIにデータを書き込み、そうでなければ指定されたLUIにのみデータを書き込んで、コンピュータにコマンド実行のステータスを返送する(ステップ902～906)。

【0047】前述したスナップショット管理プログラム406の処理において、更新先のLUIの状態が「再同期中」でアクセス位置が更新位置管理表407に記録されている場合に、更新位置管理表407の記録を削除する理由は、後述するスナップショット削除時のLUI再同期処理で実行されるデータコピー処理の無駄を無くするためである。「再同期中」の状態のLUIに対する更新において、ミラー制御プログラム415のLUIミラー機能(図6のステップ903)が有効であるため、データ更新により更新部分のLUIの内容が一致し、再同期のデータコピー処理は必要がなくなる。再同期のデータコピーは、次に説明するように、更新位置管理表407に記録された位置について実行されるため、更新により内容が一致した位置の記録は更新位置管理表407から削除する。

【0048】次に、スナップショット削除の際のデータコピーの処理について説明する。スナップショット削除の処理として前述で説明したように、スナップショットを削除する際に、二重化のペアとして定義された2つのLUIで記憶内容が一致しない部分についてオリジナルデータをミラー先LUIにコピーする処理が必要となる。この処理を行うミラー再同期機能の動作を図7に示すフローを参照して説明する。スナップショット管理プログラムのミラー再同期機能は、他の機能と並行に動作する。

【0049】(1) ミラー再同期機能は、処理を開始すると、まず、更新位置管理表407に記録があるか否かを調べ、もし、記録がなければ処理完了としてここでの処理を終了する(ステップ1000～1002)。

【0050】(2) ステップ1001で、更新位置管理

表407に記録があった場合、記録された位置の更新済みデータを1つ読み出す。この読み出しは、ディスクアレイにSCSIのREADコマンドを発行することにより行われる(ステップ1003)。

【0051】(3) ミラー再同期機能は、ステップ1003でのREADコマンドの発行に対してディスクアレイからデータが返送されてきたら、もう一度更新位置管理表407を参照し、読み出した位置の記録が削除されていないか否かを確認する。このような削除は、SCSI READコマンドを発行した直後にデータベースプログラムが同一位置の記録を更新した場合に発生する(ステップ1004)。

【0052】(4) ステップ1004の判定で、読み出した全ての位置の記録が削除されている場合、更新により新しいデータが書き込まれて2つのLUIの内容が一致しているのでデータコピーの必要がなくステップ1001の処理に戻り、読み出したデータの全部または全部の位置情報が更新位置管理表407に残っている場合、記録されている位置のデータを読み出した位置に書き戻す。LUIの状態が「再同期中」の場合、データの書き込み処理として前述で説明したように、ミラー制御プログラムは、更新をミラー先LUIにも反映させる。これにより、2つのLUIの内容を一致させることができる。(ステップ1005)。

【0053】(5) 最後に、ミラー再同期機能は、更新位置管理表407の記録を削除し、ステップ1001の処理に戻る(ステップ1006)。

【0054】前述した本発明の第1の実施形態は、コピーが必要なデータの位置をコンピュータ400内の更新位置管理表407にのみ記録することとしているため、コンピュータ400の障害により位置情報が失われる可能性がある。しかし、その場合、全てのオリジナルデータをミラー先にコピーすることにより、更新内容を失うことなく記憶内容を一致させることができる。

【0055】前述した本発明の第1の実施形態によれば、データベースプログラムがアクセスするLUIと、そのスナップショットとして使用されるミラー先LUIが異なるディスク上にあるため、バックアッププログラムとデータベースプログラムとがそれぞれ発生するディスクアクセスの負荷を分散させることができ、バックアップ中のデータベースプログラムのディスクアクセス速度の低下を軽減することができる。

【0056】すなわち、いま、2つのプログラムが領域内を均等にアクセスすると仮定し、スナップショット取得後に更新されたデータベースの割合をp、データベースプログラムが発生するディスクアクセス負荷をD、バックアッププログラムが発生するディスクアクセス負荷をBとする。すると、更新前データの保存による従来技術の方法の場合、更新されていない部分のデータは、バックアッププログラムもオリジナルデータと同じ記憶領

域を参照するため、ディスク装置 420 へのアクセス負荷は D から $(D+B(1-p))$ に増大するが、前述した本発明の第 1 の実施形態の場合、データベースアクセスとバックアップアクセスとは記憶領域を共有しないため負荷は D のままである。

【0057】また、前述した本発明の第 1 の実施形態は、スナップショットを取っていないときにデータ書き込みを二重化するが、二重化処理が、ディスクアレイ内のミラー制御プログラムで実行されているため、コンピュータは 1 回のデータ書き込みにつき、SCSI WRITE コマンドを 1 回送信するだけでよい。このため、コンピュータによるデータの二重化による従来の方法と比べ、I/O 処理のためのコンピュータの CPU 負荷と SCSI バスの通信量とを低減することができ、データベースプログラムの実行速度を向上させることができる。

【0058】さらに、前述した本発明の第 1 の実施形態で用いるディスクアレイは、スナップショット中の更新位置管理機能やスナップショット削除時の LU の再同期機能を持たないため、外部記憶装置によるデータ二重化による従来技術の方法に比較して、単純な機能のみを提供する安価なディスクアレイを用いてシステムを構成することが可能となる。

【0059】前述した本発明の第 1 の実施形態の拡張形態として、ミラー制御プログラムに機能を追加し、ミラー削除時のコンピュータ内の処理を単純化することも考えられる。この場合、指定された位置の LU 間のデータコピーを行う機能をミラー制御プログラム 415 に追加することにより、スナップショット管理プログラムは実際にデータの読み出し、書き込みを行わずにコピーすべき位置をミラー制御プログラムに指示するだけで内容を一致させることができる。その場合、図 7 のステップ 1003~1005 の代わりに、コピー位置を指示するための予め定義された SCSI コマンドを発行するステップを設ければよい。このコマンドとしては、ステップ 1003 の READ コマンドと同じアクセス位置を指定し、受信したミラー制御プログラムは、ディスクアレイ内で指定された位置のデータをミラー先 LU にコピーするコマンドであればよい。これにより、データコピー処理に伴うコンピュータの CPU 負荷や SCSI バスの通信量の増大を低減することができる。

【0060】図 8 は本発明の第 2 の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図 9 はミラー定義表の構成を示す図、図 10 は LU 状態表の構成を示す図、図 11 は Mode Sense に対して返すパラメータページの内容を説明する図であり、以下、これらの図を参照して本発明の第 2 の実施形態について説明する。図 8~図 11 において、1108、1116 はミラー定義表であり、他の符号は、図 1 の場合と同一である。この本発明の第 2 の実施形態は、前述で

説明した第 1 の実施形態と同様にスナップショットを管理するものであるが、第 1 実施形態において、コンピュータ及び外部記憶装置の双方において行われていた LU の状態管理をコンピュータのみで行い、スナップショット管理に関して外部記憶装置をステートレス化したものである。

【0061】図 8 に示す本発明の第 2 の実施形態において、ディスクアレイ 410 内のミラー制御プログラム 415 は、二重化された 2 つの LU を論理的な 1 つの LU として提供する LU 仮想化機能を有する。LU 仮想化機能で扱われる LU 間の関係は、図 9 に示すようなミラー定義表 1116 に記録される。このミラー定義表 1116 は、3 つの LU を 1 組とした LU 間の関係を表し、オリジナルデータを持つミラー元 LU、その内容を二重化したミラー先 LU、これらを論理的に 1 つの LU として扱うミラー LU を定義している。3 つの LU は、全て外部からアクセス可能である。また、ミラー元 LU とミラー先 LU との内容が一致するように保つ責任は、コンピュータ 400 で動作するスナップショット管理プログラム 406 にある。

【0062】スナップショット管理プログラム 406 は、ディスクアレイ 410 が提供する LU を仮想 LU として他のプログラムに提供する。仮想 LU は、実際の LU と同様に他のプログラムからアクセスされる。プログラムが仮想 LU にアクセスしたときにアクセス先となる実際の LU は、仮想 LU の状態により、スナップショット管理プログラムが切り替える (LU 仮想化機能)。仮想 LU のアクセス先及び状態は、図 10 に示すような LU 状態表 408 に記録される。実際にアクセス先となる LU は、読み出しの場合と書き込みの場合とで別々に定義される。LU 対応表 408 に記録される状態は、第 1 実施形態における LU 状態表と同様であるが、それ以外に「未定義」の状態があり、これはその仮想 LU に対するアクセスができない状態を表す。

【0063】更新位置管理表 407 は、図 3 により説明した第 1 の実施形態のものと同様な構造を持つが、変更された LU は、実際の LU ではなく仮想 LU が記録される。また、スナップショット管理プログラムの更新監視機能、更新位置管理機能、ミラー再同期機能がそれぞれ監視、管理、アクセスする対象 LU も仮想 LU である。各機能は、第 1 の実施形態の場合と同様に仮想 LU の状態に応じて有効化、無効化がなされる。「未定義」の LU に対して、全ての機能は無効化される。

【0064】また、スナップショット管理プログラム 406 は、ディスクアレイ内のミラー定義表 1116 と同一内容のミラー定義表 1108 を持つ。このミラー定義表は、スナップショット管理プログラムの起動時に、全てのアクセス可能な LU に対して SCSI Mode Sense コマンドを発行することにより、ディスクアレイ 410 からミラー定義表 1116 の情報を得て作成される。デ

ィスクアレイ410は、Mode Senseに対して図11に示すような構造を持つ4バイトのパラメータページを返送し、各LUIについてミラー定義表1116の内容を知らせる。

【0065】このパラメータページは、先頭の1バイト目にはページコード0が入り、ModeSenseで指定されたLUIがミラーLUIであれば2バイト目には“0”、そうでなければ“1”が入る。また、種別が“0”のとき、3バイト目、4バイト目にはそれぞれミラー元LUI、ミラー先LUIが入る。ミラー定義表1108を作成する場合、種別が“0”のLUIに対してエントリを作成し、ミラー元LUI、ミラー先LUIを記録する。このミラー定義表1108の作成は、Mode Senseを用いた自動作成である必要はなく、ユーザが手動で定義した内容ファイルをスナップショット管理プログラム406が起動時に読み込んで作成してもよい。

【0066】初期状態において、LUI状態表408のデータベースプログラム404がアクセスする仮想LUIの状態は「同期」、READ先及びWRITE先LUIはいずれもミラー定義表1108のミラーLUIが記録され、バックアッププログラム405がアクセスする仮想LUIの状態は「未定義」が記録されている。

【0067】次に、前述のように構成される本発明の第2の実施形態におけるコンピュータ上のプログラムのディスクアクセスについて説明する。

【0068】コンピュータ上のプログラムからのディスクアクセスは、スナップショット管理プログラム406が提供する仮想LUIに対して要求される。スナップショット管理プログラム406は、要求を受け取ると、LUI状態表408を参照し、指定された仮想LUIの状態が「未定義」でないか否かを確認し、「未定義」であればアクセスを拒否する。

【0069】次に、スナップショット管理プログラム406は、指定された仮想LUIのREAD先LUI、WRITE先LUIをLUI状態表408で調べ、発行するコマンドのアクセス先となる実際のLUIを決定する。例えば、図10に示す状態において、仮想LUI0にアクセスすると、スナップショット管理プログラム406は、読み出し要求、書き込み要求共に、LUI0をターゲットとするSCSIコマンドを発行する。また、仮想LUI4に読み出し、書き込みを要求すると、スナップショット管理プログラム406は、それぞれLUI7、LUI6にコマンドを発行する。スナップショット管理プログラム406は、LUI状態表を書き換えることにより、各プログラムがアクセスする仮想LUIを固定したままで、アクセス先のLUIを切り替えることができる。

【0070】アクセスコマンドを受信したディスクアレイのミラー制御プログラム415は、ミラー定義表1116を参照して対応するLUIへのアクセスを実行する。例えば、図9に示すように定義されているLUI0に対す

るアクセスを受信すると、LUIミラー制御プログラム415は、ミラー定義表を参照してLUI0が2つのLUIからなるミラーLUIであると判断し、更新ではLUIミラー機能によりLUI1とLUI2に同じ内容を書き込み、読み出しではLUI1またはLUI2からデータを読み出す。LUI1やLUI2にアクセスがあった場合、これらはミラーLUIではないとして、それぞれLUI1またはLUI2にのみアクセスする。

【0071】次に、本発明の第2の実施形態におけるスナップショットの取得、削除について説明する。前述した本発明の第1の実施形態は、スナップショットの取得・削除の際に、ディスクアレイに対してMode Select コマンドを発行してスナップショットを管理したが、第2の実施形態は、LUI状態表408に記録された仮想LUIのアクセス先LUIを変更することにより、スナップショットを管理する。

【0072】スナップショットの取得前、LUI状態表408には、READ先LUI、WRITE先LUI共にミラー定義表のミラーLUIが記録されている。スナップショット取得時、スナップショット管理プログラム406は、これらをミラー定義表に記録されているミラー元LUIに書き換える。また、スナップショット管理プログラム406は、バックアッププログラム405がアクセスするスナップショットの仮想LUIのREAD先、WRITE先LUIをミラー先LUIとし、状態を「スナップショット中」とする。

【0073】例えば、図9及び図10に示す状態で、仮想LUI0のスナップショットを取得するとき、スナップショット管理プログラム406は、現在のアクセス先であるLUI0のミラー元LUIが、図9のミラー定義表を参照することによりLUI1であることが判るので、仮想LUI0のREAD先LUI、WRITE先LUIをLUI1に書き換え、状態を「スナップショット中」とする。また、スナップショット管理プログラム406は、仮想LUI0のスナップショットとしてアクセスされる仮想LUI1のREAD先LUI、WRITE先LUIをLUI0のミラー先LUIであるLUI2に書き換え、状態を「スナップショット中」とする。これにより、仮想LUI1へのアクセスが可能となり、また、仮想LUI0、1の更新は、互いに二重化されないため、オリジナルデータ及びスナップショットのアクセスにそれぞれの仮想LUIを利用できるようになる。

【0074】スナップショット管理プログラム406は、スナップショットを削除する場合、LUI状態表408に記録されたスナップショットの仮想LUIの状態を「未定義」とし、オリジナルデータの仮想LUIの状態を「再同期中」に変更し、また、オリジナルデータの仮想LUIのWRITE先LUIをミラー定義表に記録されているミラーLUIとする。例えば、図9及び図10に示す状態で、仮想LUI2のスナップショットである仮想LUI3

を削除するとき、スナップショット管理プログラム406は、仮想LU3の状態を「未定義」とし、また、仮想LU2の現在のWRITE先LUであるLU4のミラーLUが図9のミラー定義表を参照することによりLU3であることがわかるので、WRITE先LUをLU4からLU3に書き換えると共に、状態を「再同期中」とする。これにより、スナップショットである仮想LU3へのプログラムからのアクセスが不可能となり、また、ミラー元LUとミラー先LUとの内容を一致させるデータコピー処理の際の、オリジナルデータの仮想LUからの読み出しはミラー元LU、オリジナルデータの仮想LUへの書き戻しはミラーLUがそれぞれアクセス先となる。

【0075】データコピー処理が完了した後、スナップショット管理プログラム406は、オリジナルデータとしてアクセスされる仮想LUのREAD先LUを元のミラーLUに変更する。例えば、図9及び図10の状態では、「再同期中」の状態にある仮想LU4のスナップショット削除を完了するために、スナップショット管理プログラム406は、図9のミラー定義表から現在のREAD先LUであるLU7のミラーLUがLU6であることが判るので、READ先LUをLU6に書き換え、状態を「同期」とする。これにより、スナップショットの削除が完了する。

【0076】前述した本発明の第2の実施形態によれば、第1の実施形態で説明した効果に加え、外部記憶装置をステートレス化したことにより障害時の回復処理の単純化を図ることができる。例えば、LUの状態を変更する瞬間にコンピュータが障害を起こした場合、第1の実施形態の場合、回復処理においてコンピュータのLU状態表に記録されている各LUの状態とディスクアレイのミラー管理表に記録されている各LUの状態を一致させる必要があるが、第2の実施形態の場合、ディスクアレイがステートレスであるため、コンピュータの状態のみを考慮した処理を行うことができる。

【0077】図12は本発明の第3の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図であり、以下、これについて説明する。この本発明の第3の実施形態は、データベースプログラムとバックアッププログラムとがそれぞれ別のコンピュータで動作する。この実施形態は、2つのコンピュータがLANで接続され、各コンピュータとディスクアレイとがファイバチャネルを用いたStorage Area Network(SAN)により接続され、また、バックアップ開始・終了を指示する管理コンソールとして別のコンピュータがLANで接続されて構成されている。

【0078】図12に示す第3の実施形態による計算機システムは、データベースサーバ1510、バックアップサーバ1520、管理コンソール1540が、それぞれLANインタフェース1513、1523、1541

を介してLAN1500により接続されて構成される。また、データベースサーバ1510とバックアップサーバ1520とは、ファイバチャネルインタフェース1514、1524を持ち、ファイバチャネルケーブル1531~1533とファイバチャネルスイッチ1530を介してディスクアレイ1550に接続されている。データベースサーバ及びバックアップサーバは、ファイバチャネルネットワークを経由してディスクアレイ1550内のデータにアクセスすることができる。ディスクアレイ1550の機能は、通信がファイバチャネル上に定義されるSCSIプロトコルに置き換えられた以外、第2の実施形態の場合と同一である。

【0079】図12において、データベースサーバ1510は、データベースプログラム1515、スナップショット管理プログラム1517、通信プログラム1516を実行し、バックアップサーバ1520は、バックアッププログラム1526及び通信プログラム1525を実行する。通信プログラム1516は、管理コンソールとLANで通信することによりバックアップ処理に関する指示受け取り、その内容に基づいてデータベースプログラム、スナップショット管理プログラムを呼び出す。同様に、通信プログラム1525は、管理コンソール1540からの指示に基づいてバックアッププログラム1526を呼び出す。データベースプログラム1515及びスナップショット管理プログラム1517の機能は、通信プログラムから指示を受けることができる点を除き第2の実施形態の場合と同様である。バックアッププログラム1526は、バックアップサーバ1520上にスナップショット管理プログラムがないため、仮想LUではなくディスクアレイが提供する実際のLUを指定してアクセスを行う。

【0080】管理コンソール1540は、バックアップ指示プログラム1542を実行し、LANインタフェース1541とLAN1500とを経由してデータベースサーバ1510及びバックアップサーバ1520上の通信プログラム1516、1525に指示を送る。

【0081】次に、前述のように構成される本発明の第3の実施形態における管理コンソールとサーバとの通信について説明する。

【0082】管理コンソール1540でユーザがバックアップの指示を出すと、バックアップ指示プログラム1542は、データベースサーバ1510の通信プログラム1516に対し、スナップショットを取得するように指示を行う。この指示を受信した通信プログラム1516は、データベースプログラム1515に対して、更新動作を一時停止するよう指示し、スナップショット管理プログラム1517にスナップショットの取得を指示する。スナップショット取得後、通信プログラム1516は、再びデータベースプログラムに指示を出してデータベースの更新を再開させ、スナップショット取得完了を

バックアップ指示プログラム1542に報告する。

【0083】スナップショット取得完了の報告を受けたバックアップ指示プログラム1542は、次にバックアップサーバ1520の通信プログラム1525に対し、バックアップを開始するように指示を行う。この指示を受けた通信プログラム1525は、バックアッププログラム1526にバックアップ開始を指示し、バックアップを行わせる。バックアップ完了後、通信プログラム1526は、バックアップ完了をバックアップ指示プログラム1542に報告する。

【0084】バックアップ完了の報告を受けたバックアップ指示プログラム1542は、通信プログラム1516に対してスナップショットの削除を指示する。この指示を受けた通信プログラム1516は、スナップショット管理プログラム1517に対して、スナップショットを削除するように指示を行う。スナップショット削除後、通信プログラム1516は、スナップショット削除完了をバックアップ指示プログラム1542に報告し、これにより、データベースのオンラインバックアップが完了する。

【0085】前述した本発明の第3の実施形態によれば、第2の実施形態で説明した効果に加え、バックアップ中のデータベースサーバのCPU負荷の増大を軽減することができる。第2の実施形態は、バックアッププログラムとデータベースプログラムとが同一コンピュータ上で動作するため、バックアップ中のデータ転送処理にともなうCPU負荷増大により、データベースプログラムの実行速度が低下していたが、前述した本発明第3の実施形態は、バックアッププログラムがデータベースサーバとは別のコンピュータで動作するため、データ転送処理の負荷がデータベースサーバのCPU負荷に影響を与えることがなく、バックアップ中のデータベースプログラムの実行速度の低下を防止することができる。

【0086】また、前述した本発明の第3の実施形態によれば、バックアップサーバとディスクアレイとの間がファイバチャネルで接続されているため、高速なデータ転送による迅速なバックアップを実現することができる。

【0087】図13は本発明の第4の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図14はミラー定義表の構成を示す図であり、以下、これらの図を参照して第4の実施形態について説明する。この本発明の第4の実施形態は、更新前データの保存による従来技術のスナップショット管理方法から、前述した本発明の第1、第2の実施形態によるスナップショット管理方法に移行する場合に、従来のスナップショット管理プログラムの変更を最小限にとどめつつ本発明を適用し、スナップショット管理プログラムの大幅な変更を不要としたものである。

【0088】ここで説明する本発明の第4の実施形態

は、記憶内容が2つの異なる記憶媒体に二重化されるミラーLUにオリジナルデータを記憶し、スナップショット取得時点以後に更新が要求された場合に、更新前のデータを別領域に保存してからオリジナルデータを更新するものである。スナップショットは、スナップショット管理プログラムにより1つの仮想的なLUとして表されるが、更新されていない部分は、オリジナルデータを、更新された部分は、別領域に保存した更新前データを参照することによりアクセスされる。この第4の実施形態は、スナップショット管理プログラムの基本的な動作としては、更新前データの保存による従来の方と同一であるが、ディスクアレイの支援機能と連携するために後述するアクセス先振り分け機能が追加されている。

【0089】図13に示す本発明の第4の実施形態におけるハードウェアの構成及びデータベースプログラム、バックアッププログラムの機能は前述で説明した本発明の第1、第2の実施形態と同様である。図13において、ディスクアレイ410内で動作するミラー制御プログラム415は、前述した本発明の第2の実施形態の場合と同様に、LU間の関係を定義したミラー定義表1116を有している。ここでのミラー定義表1116は、その構造を図14に示しているが、前述で説明した本発明の第2の実施形態における図9に示したミラー定義表とは異なり、ミラーLUとミラー先LUとのみが定義される。第2の実施形態の場合と同様に、ミラーLUは、二重化された2つのLUを仮想的に1つのLUとして表したものであり、ミラー先LUは、ミラーLUを構成する2つのLUの1つである。図14に示す定義の場合、LU0はLU1に二重化され、ミラー制御プログラム415は、LU0に対するアクセスコマンドを受け取ると、読み出しであればLU0またはLU1からデータを読み出し、書き込みであれば同じ内容をLU0とLU1との両方に書き込む。ミラー先LUであるLU1に対するアクセスは、LU1にのみアクセスし、LU0にアクセスすることはできない。

【0090】また、ミラー定義表1116でミラー先LUとして記録されているLUは読み出し専用であり、ミラー先LUに対する書き込み要求は、ミラー制御プログラム415により拒否される。これにより、ミラー元LUとミラー先LUとの内容を常に一致させることができる。ミラーLUから読み出すデータとミラー先LUから読み出すデータとの内容は同一であるが、ミラー先LUにアクセスした場合、アクセス先のLUが1つに限定される点が異なる。データベースプログラム404等のコンピュータ上のアプリケーションプログラムは、常にミラーLUに対してアクセスを行う。ミラー先LUは、後述するスナップショット管理プログラム406のアクセス先振り分け機能で使用するために提供される。

【0091】スナップショット管理プログラム406は、オリジナルデータを参照するために従来の方の場合

合にミラーＬＵに対して発行されるアクセスコマンドの一部を、ミラー先ＬＵに振り分けるアクセス先振り分け機能を有している。オリジナルデータのミラーＬＵに対応するミラー先ＬＵは、ミラー定義表１１０８を参照することにより得られる。ミラー定義表１１０８は、ディスクアレイ４１０内のミラー定義表１１１６と同一の内容を持ち、第２の実施形態の場合と同様に、スナップショット管理プログラム４０６のMode Senseによるディスクアレイ４１０からの情報取得や、ユーザの手動設定によって作成される。

【００９２】次に、スナップショット管理プログラム４０６のアクセス先振り分け機能について説明する。

【００９３】オリジナルデータに対するアクセスは２種類に分けられる。その１つは、データベースプログラムからのオリジナルデータへのアクセスである。また、もう１つは、バックアッププログラムからのスナップショットへのアクセスであって、アクセス先のデータがスナップショット取得時点以後に更新されていない場合のアクセスであり、この場合、スナップショットが、オリジナルと記憶領域を共有しているのでアクセス先はオリジナルデータとなる。

【００９４】従来技術の方法をそのまま適用すると、前述のどちらのアクセスでも、オリジナルデータのアクセス用に提供されるミラーＬＵに対してアクセスコマンドを発行することになるが、本発明の第４の実施形態におけるスナップショット管理プログラム４０６は、後者のアクセス先をミラー先ＬＵに変更する。

【００９５】これにより、バックアッププログラム４０５から発生するオリジナルデータへの全ての読み出し要求は、そのアクセス先がミラー先ＬＵに限定され、従って、そのアクセス負荷がミラー先ＬＵが割り当てられたディスクに限定されることになる。

【００９６】一方、データベースプログラム４０４から発生する読み出し要求のアクセス先はミラーＬＵであり、二重化された２つのＬＵのどちらをも使用することができる。従って、この読み出し要求のアクセス先として、負荷の低いディスクを使用することにより、バックアッププログラムから発生するオリジナルデータに対する読み出しの負荷が高い場合でも、データベースプログラムのディスクアクセス速度が低下することはない。例えば、図１４に示すミラー定義表のＬＵ０、ＬＵ１が、それぞれ、図１３のディスク装置４２０、４２１に割り当てられている場合に、ＬＵ０のスナップショットを取得すると、バックアッププログラムからオリジナルデータへのアクセス要求は、全てＬＵ１をアクセス先とするためディスク４２１のみを使用する。データベースプログラム４０４から発生するアクセス要求のアクセス先はＬＵ０であるため、ディスク４２０、４２１のどちらでも使用することができる。このため、ミラー制御プログラム４１５が負荷の低いディスク４２０をアクセス先と

して選択すれば、バックアップによるディスクアクセス負荷が高い場合でも、データベースプログラムのディスク読み出し性能を低下させることがない。

【００９７】前述した本発明の第４の実施形態によれば、第２の実施形態の場合の効果に加え、従来技術のスナップショット管理プログラムを流用することができ、スナップショット管理プログラムの開発期間を短縮することができる。すなわち、第４の実施形態によれば、従来技術のスナップショット管理プログラムに、アクセス先振り分け機能を追加することにより、ディスクアレイの支援機能を利用するスナップショット管理プログラムを容易に作成することができる。

【００９８】前述した本発明の各実施形態によれば、オリジナルデータのアクセスに使用される記憶領域とスナップショットのアクセスに使用される記憶領域とが異なる記憶媒体上にあるため、バックアップ中の特定媒体への負荷の集中を避けることが可能となり、また、バックアップをオンライン処理とは別のコンピュータ上で実行することができ、これにより、バックアップ中のＣＰＵ負荷の増大を軽減することができる。

【００９９】また、本発明の各実施形態によれば、データ書き込みの二重化を外部記憶装置において行っているため、データ書き込みの二重化処理に伴うコンピュータのＣＰＵ負荷やデータ通信量の増加を生じさせることがなく、さらに、スナップショット取得中の更新位置の管理機能や、スナップショット削除の際に二重化された記憶内容の一致を保証する機能を持たない外部記憶装置を用いることができ、これにより、安価な計算機システムを構成することができる。

【０１００】

【発明の効果】以上説明したように本発明によれば、バックアップ中の特定媒体への負荷の集中を避けることが可能となり、また、データ書き込みの二重化処理に伴うコンピュータのＣＰＵ負荷やデータ通信量の増加を生じさせることを防止することができる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。

【図２】ミラー管理表の構成を示す図である。

【図３】更新位置管理表の構成を示す図である。

【図４】Mode Select及びMode Senseで使用するパラメータページの内容を説明する図である。

【図５】スナップショット管理プログラムのデータ書き込み時の動作を説明するフローチャートである。

【図６】コンピュータからWRITEコマンドを受信したミラー制御プログラムの動作を説明するフローチャートである。

【図７】スナップショット管理プログラムのミラー再同期機能の動作を説明するフローチャートである。

【図8】本発明の第2の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。

【図9】ミラー定義表の構成を示す図である。

【図10】LU状態表の構成を示す図である。

【図11】ModeSense に対して返すパラメータページの内容を説明する図である。

【図12】本発明の第3の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。

【図13】本発明の第4の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。

【図14】ミラー定義表の構成を示す図である。

【符号の説明】

400 コンピュータ

401、411 CPU

401、412 メモリ

403、413 SCSI インタフェース

404 データベースプログラム

405 バックアッププログラム

406 スナップショットプログラム

407 更新位置管理表

408 LU状態表

409 SCSI バス

410 ディスクアレイ

414 ディスクアレイ

415 ミラー制御プログラム

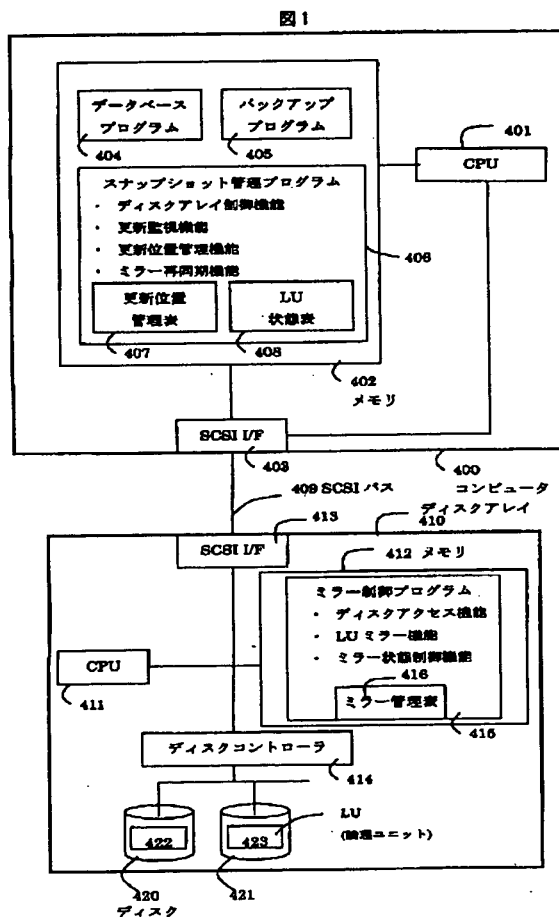
416 ミラー管理表

420、421 ディスク装置

422、423 LU

1108、1116 ミラー定義表

【図1】



【図2】

LU	ミラー先 LU	状態
1	4	同期
2	5	スナップショット中
3	6	再同期中

【図11】

ページコード 00h
種別
ミラー元 LU
ミラー先 LU

【図3】

LU	更新部開始 LBA	更新部終了 LBA
2	0	15
2	256	511
3	8	31

【図4】

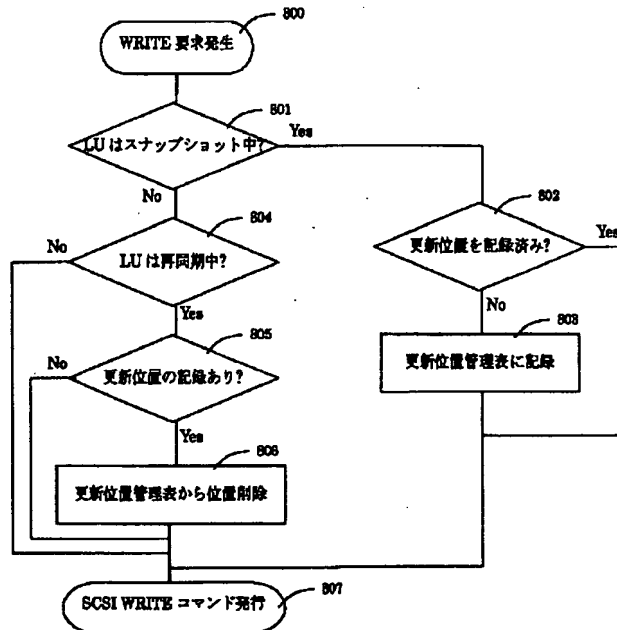
ページコード 00h
状態

【図14】

ミラーLU	ミラー先 LU
0	3
1	4
2	5

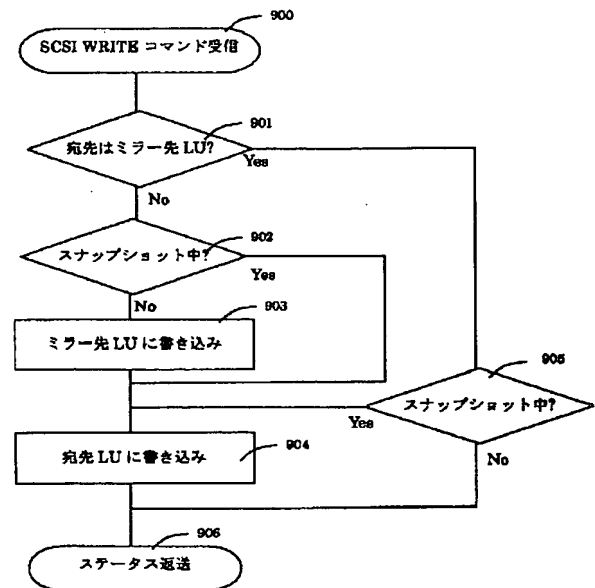
【図 5】

図 5



【図 6】

図 6



【図 8】

【図 7】

図 7

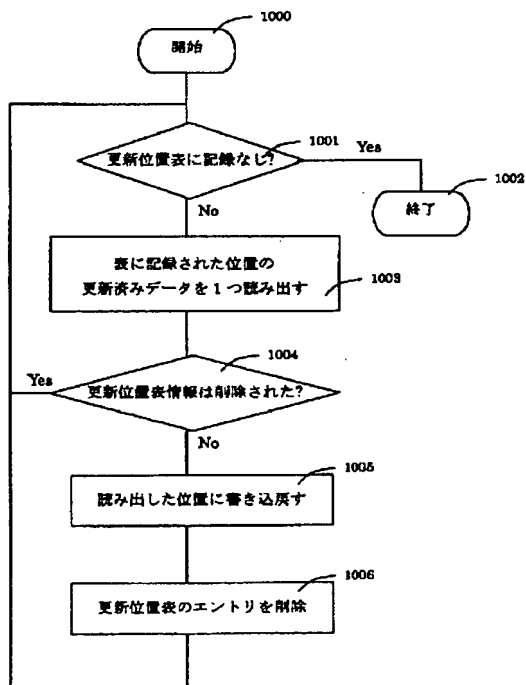
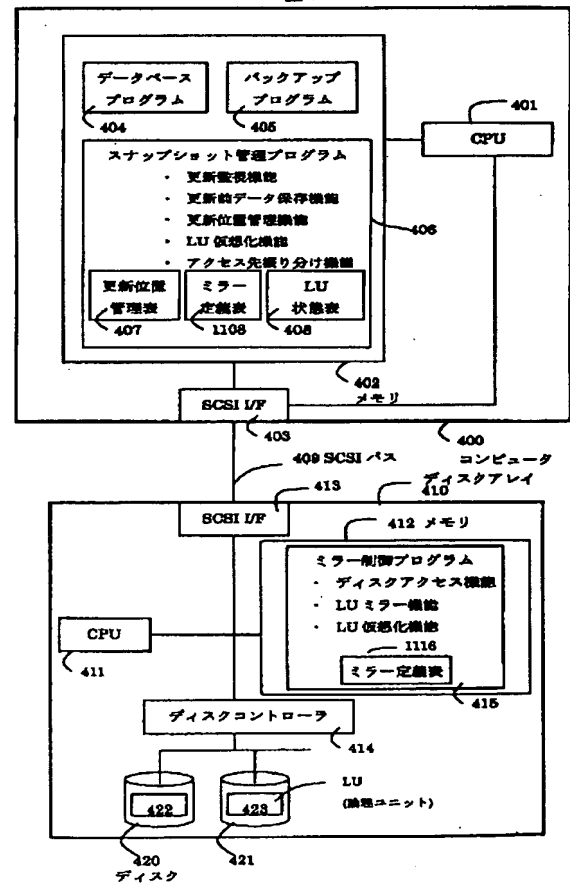
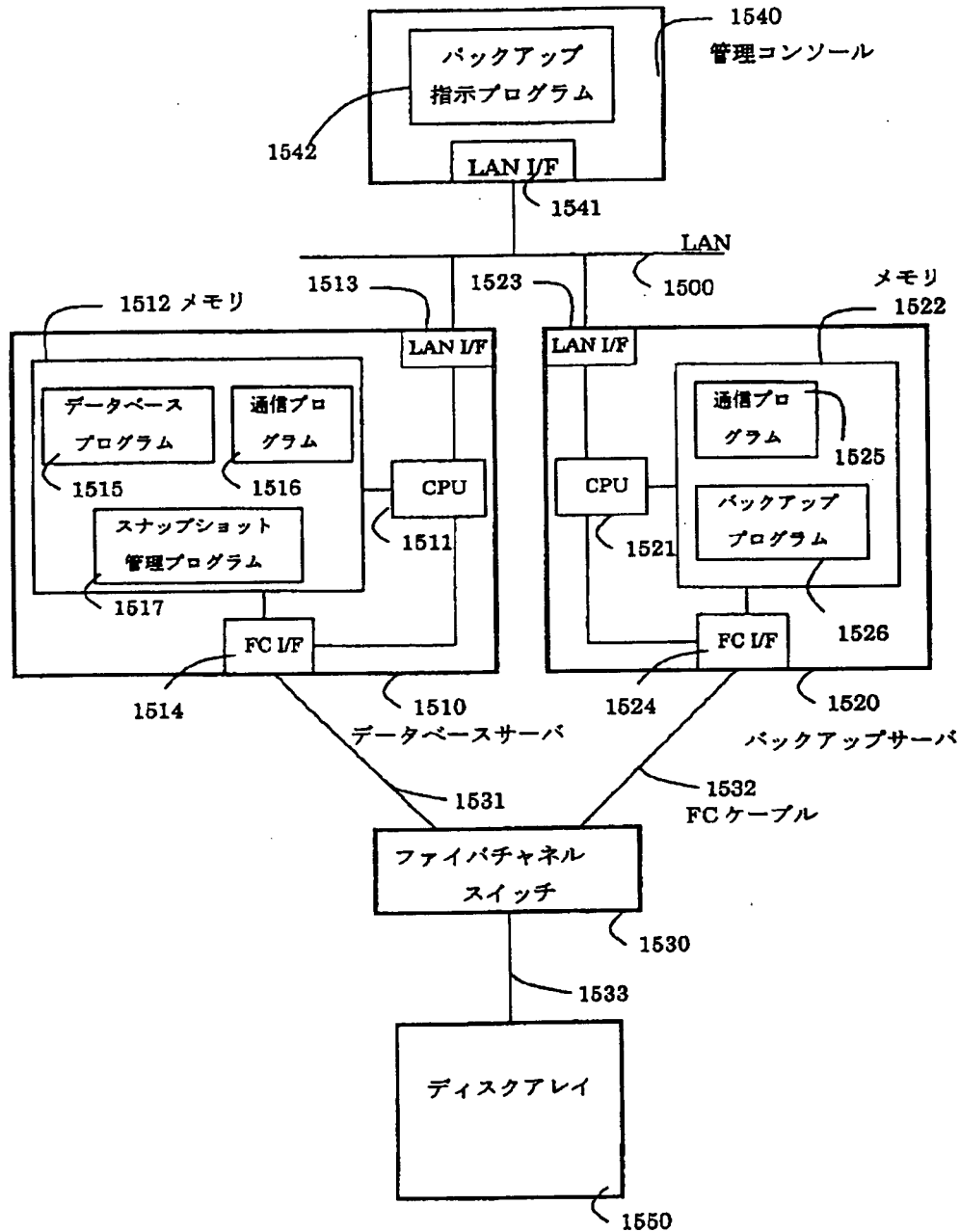


図 8



【図12】

図12



フロントページの続き

(72) 発明者 村岡 健司
神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内
(72) 発明者 大枝 高
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内

(72) 発明者 八木沢 育哉
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内
F ターム(参考) 5B018 GA04 HA03 MA12
5B065 BA01 CA30 EA02 EA12 EA34
5B083 AA08 AA09 BB01 BB03 CC04
CD11 EE08
5D044 AB03 BC01 CC04 DE03 DE12
DE72 DE92 HL02